## SUPPLEMENTAL INFORMATION

### Bell and Leggett-Garg Inequalities

The CHSH correlator, designed by Clauser, Horne, Shimony, and Holt[1] as a refinement of the Bell inequality[2], provides a quantitative bound on classical hidden variable theories using correlated measurements between two spatially separated qubits. The correlator combines four different experimental configurations because it can be difficult to tell the difference between potentially classical (un-entangled) qubits and an entangled state in only one basis. With superconducting qubits, the measurement basis for each qubit is set using qubit rotations to map the desired state onto the ground ($|0\rangle$) and excited ($|1\rangle$) states of the system. For measurement rotations $a$ (qubit 1) and $b$ (qubit 2), shown in Fig. 1(a), the correlation amplitude is given by

$$E(a,b) = P(00) - P(10) - P(01) + P(11), \tag{1}$$

where P(00) is the probability both qubits are in the ground state. Given this equation we can see that both the Bell state $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ and the prepared state $|00\rangle$ will have a correlation amplitude of 1 if $a = b = 0$. The difference only becomes clear when the detector angle of one qubit is rotated relative to the other. The behavior of $E(a,b)$ vs detector rotation, described here as $\theta = a - b$, is shown in Fig. 1(b) for both the classical and quantum case. If the two objects can be described separately, then $E$ is only a linearly dependent on $\theta$. If the two objects are entangled, then $E$ is has a sinusoidal dependence on $\theta$ with the maximum difference occurring at $\theta = \pi/4$.

To initially characterize the system we conducted a traditional CHSH experiment using the central Bell qubits ($\beta_{1,2}$). The relative measurement angles for each qubit were held fixed such that $a' = a + \pi/2$ and $b' = b + \pi/2$. We then varied $\theta = a - b$ from 0 to $\pi$, and measured each individual correlator as well as the sum given by

$$\text{CHSH} = E(a,b) + E(a',b) + E(a,b') - E(a',b'). \tag{2}$$

For any two classical states measured at these angles, we should see a linear dependence of $E(\theta)$ and a bound on the the CHSH correlator of $|CHSH| \leq 2$. Alternately, if the two qubits are in a maximally entangled Bell state, we should see sinusoidal behavior for $E(\theta)$ and a maximum CHSH value of $2\sqrt{2}$. The data, shown in Fig. 1(c), display the expected sinusoidal dependence for each individual term, with a maximum CHSH amplitude near $\theta = \pi/4$. While this data shows a robust violation of the classical bound, it fails to reach the theoretical maxium bound of $2\sqrt{2}$.

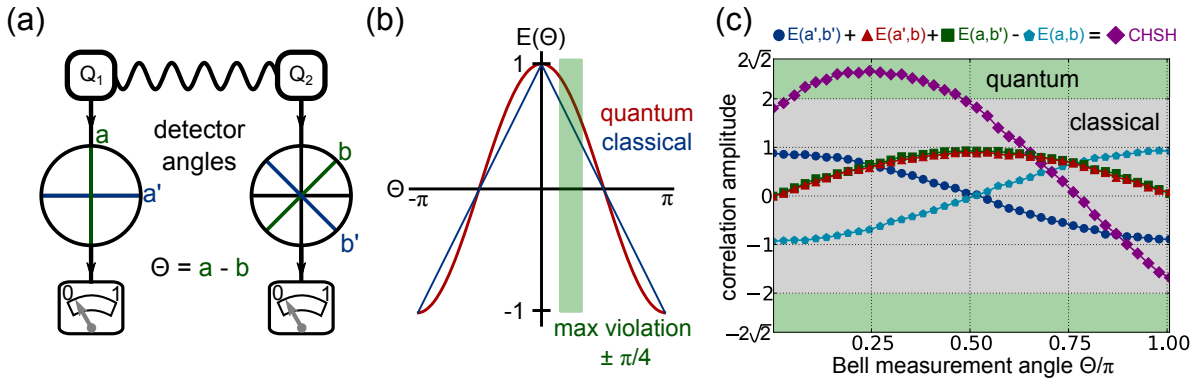


Figure 1. Entanglement and the CHSH corralator. (a) Schematic of a typical CHSH experiment. Two qubits are prepared in one of the four possible Bell states and measurements are conducted at various measurement angles $a$, $b$, $a' = a + \pi/2$, and $b' = b + \pi/2$. (b) Dependence of correlation amplitude (Eq. 1) on detector angle difference $\theta = a - b$. The difference between quantum and classical correlation is maximized at $\theta = \pi/4$. (c) Measurement of CHSH correlator using the central Bell qubits prepared in a $|\phi^+\rangle$ Bell state, displaying the sinusoidal dependence on the relative measurement angle $\theta$. This experiment avoids common sampling loopholes with high detector efficiency, but is subject to the locality loophole due to the small spatial separation of adjacent superconducting qubits on a chip.

The maximum CHSH amplitude of $\sim 2.5$ we see here is due to experimental imperfections which will be discussed later. This CHSH experiment provides the framework for the BLGI, as well as a benchmark for the maxium violation we should expect at the weakest measurement angles.

A complementary test of quantum mechanics is the LGI, which is similar to a Bell inequality but involves measurements separated in time rather than in space. Classical theories of measurement assume that the system is always in a definite state, and that an ideal measurement will not change the state of the system. In contrast, if one were to measure a quantum state in an orthogonal measurement basis, the act of measurement would project that object onto an eigenstate of the new basis. To distinguish one kind of system from the other, measurements are conducted in different bases at different times. For measurements conducted at times $t_1 < t_2 < t_3$, we can construct correlators analogous to Eq. 1 but for different measurements of the same qubit,

$$E(t_i, t_j) = P(00) - P(10) - P(01) + P(11). \tag{3}$$

The inequality was originally composed of three distinct experiments. In the first experiment, the system is measured projectively at time $t_1$, followed by a final projective measurement at time $t_3$. A second experiment is then carried out where an intermediate measurement in a different basis is

conducted at time $t_2$ instead of time $t_3$. The third experiment consists of only the measurements at times $t_2$ and $t_3$. The LGI is then given by

$$-3 \leq E(t_1, t_2) + E(t_2, t_3) - E(t_1, t_3) \leq 1 \tag{4}$$

where $E(t_1, t_3)$ is the experiment in which no measurement is performed at time $t_2$. Further details for LGIs can be found in the review article by Emary *et al.* [3].

The weak measurement techniques discussed in the main text were created to avoid the possibility of a "clumsy" measurement loophole [4]. When sequential measurements are performed on the same system, it is impossible to ensure that LGI violations are not due to overly invasive measurements perturbing the system in an unknown way. To minimize the effect of measurement, most LGIs replace the measurement at time $t_1$ with preparation of a known state, and the measurement at time $t_2$ with a null [5] or weak [6, 7] measurement. These weak measurements [6, 7] minimize back action on the system, while still extracting enough information to identify its state. Using this technique, all the statistics of the LGI can be measured by conducting all three "measurements" in a single experimental configuration. To construct the Bell-Leggett-Garg inequality we combined a traditional CHSH experiment with this weak measurement technique. This allows us to measure all four terms of the CHSH correlator simultaneously in a single experiment.

**BLGI Algorithm Assumptions and Loopholes**

The fundamental assumptions of the hybrid Bell-Leggett-Garg inequality are those of *local realism*, which are familiar from the Bell inequalities:

(i) If an object has several distinguishable physical states $\lambda$, then at any given time it occupies only one of them.

(ii) A measurement performed on one object of a spatially-separated pair cannot disturb the second object.

(iii) Measured results are determined causally by prior events.

Note that only assumption (ii) differs from the notion of *macrorealism* used in Leggett-Garg inequalities: it is weakened here to permit *local* invasiveness for sequential measurements in time made on the same object, while still forbidding spatially *remote* measurements from influencing

each other. Note that the assumed physical state $\lambda$ may be related to the quantum state, or may be a collection of more refined (but unspecified) hidden variables.

To these core assumptions we must append one more to permit noisy (i.e., realistic) detectors:

(iv) Unbiased noisy detectors produce results that are correlated with the true object state $\lambda$ on average.

This assumption can be understood as follows. The object state $\lambda$ ideally determines each measurable property $A(\lambda)$, but a physical detector (and environment) that interacts with the system will also have a distinct physical state $\xi$ that may fluctuate noisily between realizations (e.g., from the coupling procedure). In such a case the detector will report a correspondingly fluctuating signal $\alpha(\xi)$ according to some response probability $P_A(\xi|\lambda)$ for obtaining the detector state $\xi$ given each definite system state $\lambda$. For any sensible detector, these response probabilities will be fixed by the systematic and repeatable coupling procedure (such as our ancilla measurement circuit). To calibrate such a detector, we must then assume that averaging over many realizations of the detector noise will faithfully reflect information about each *prepared* system state $\lambda$ (even if that state ultimately changes for subsequent measurements due to the coupling):

$$\textstyle\sum_\xi \alpha(\xi)\, P_A(\xi|\lambda) = A(\lambda). \tag{5}$$

Importantly, this equality formally states only what is usually assumed for an unbiased laboratory detector: that one can recover a meaningful system value $A(\lambda)$ by averaging away any detector noise.

Now consider the Bell-Leggett-Garg correlation. A correlated pair of objects with the joint state $\lambda$ is sampled from an ensemble with the distribution $P(\lambda)$. (In our experiment, we prepare two qubits in a Bell state.) At a later time each object ($k = 1, 2$) is coupled to a detector (an ancilla qubit) that outputs a noisy signal $\alpha_k$ calibrated to measure the bounded property $A_k(\lambda) \in [-1, 1]$ on average (the $Z$ operator for each Bell qubit). The noisy signal $\alpha_k$ generally has an expanded range of values that can lie outside the range $[-1, 1]$ (in our case $\alpha_k \approx \pm 1/\sin\phi$); however, for each $\lambda$ the realizations of the output signal will average to the correct bounded value by assumption (iv). (We verify this assumption with the ancilla calibration measurements using definite preparations of $0$ or $1$ on the Bell qubits.) Finally, each object is measured with a second detector that outputs a signal $b_k$ for a similarly bounded property $B_k(\zeta) \in [-1, 1]$ (we read out the qubits directly to obtain $b_k = \pm 1$). From these four measured numbers, we then compute the

CHSH-like correlator as a single number for each preparation

$$C = -\alpha_1\alpha_2 - \alpha_1 b_2 + b_1\alpha_2 - b_1 b_2. \tag{6}$$

The expanded ranges of the noisy signals $\alpha_k$ generally produce a similarly expanded range for the correlator $C$ for each preparation. Nevertheless, averaging $C$ over many realizations of the detector noise $\xi_{A_k}$ and $\xi_{B_k}$ and system states $\lambda$ will produce

$$\begin{aligned}\langle C\rangle &= \textstyle\sum_\lambda \sum_{\substack{\xi_{A_1},\xi_{B_1}\\ \xi_{A_2},\xi_{B_2}}} C\, P(\xi_{A_1},\xi_{B_1}|\lambda)P(\xi_{A_2},\xi_{B_2}|\lambda)P(\lambda),\\ &= \textstyle\sum_\lambda \big[ - A_1(\lambda)A_2(\lambda) - A_1(\lambda)B_2'(\lambda) + B_1'(\lambda)A_2(\lambda) - B_1'(\lambda)B_2'(\lambda)\big]\, P(\lambda),\end{aligned} \tag{7}$$

with $A_k(\lambda) = \sum_{\xi_{A_k},\xi_{B_k}} \alpha_k(\xi_{A_k})\, P(\xi_{A_k},\xi_{B_k}|\lambda)$ and $B_k'(\lambda) = \sum_{\xi_{A_k},\xi_{B_k}} b_k(\xi_{B_k})\, P(\xi_{A_k},\xi_{B_k}|\lambda)$, since postulate (ii) causes the joint distribution of the detector states to factor: $P(\xi_{A_1},\xi_{B_1},\xi_{A_2},\xi_{B_2}|\lambda) = P(\xi_{A_1},\xi_{B_1}|\lambda)P(\xi_{A_2},\xi_{B_2}|\lambda)$ in the same way as for a Bell inequality. From the postulates (i), (iii), and (iv), the averages $A_k(\lambda)$ and $B_k'(\lambda)$ are then bounded to the range $[-1,1]$. Therefore, for each $\lambda$ the sum of the bounded averages in Eq. (7) must itself be bounded by $[-2,2]$. Averaging this bounded result over $P(\lambda)$ produces the expected BLGI

$$-2 \le \langle C\rangle \le 2. \tag{8}$$

Importantly, the joint probability $P(\xi_{A_k},\xi_{B_k}|\lambda) = P(\xi_{A_k}|\lambda)P(\xi_{B_k}|\lambda,\xi_{A_k})$ for each qubit $k$ admits the dependence of the $B_k'$ measurement on an *invasive* $A_k$ measurement that can alter the physical state $\lambda$. Despite any randomization of the results $b_k(\xi_{B_k})$ caused by such local invasiveness, however, the perturbed averages $B_k'(\lambda)$ must still lie in the range $[-1,1]$ since each $b_k = \pm 1$ by construction. This allowance for *locally* invasive measurements in the BLGI is what avoids the clumsiness loophole [4] of the usual LGI. The fact that the entire correlator $C$ is computed for every realization in the same experimental configuration is what avoids any variant of the disjoint sampling loophole [8] for the usual Bell and LGIs (such as from systemic bugs in the preparation software).

There are, however, two notable ways that our derivation of the BLGI in Eq. (8) could fail. First, the assumptions (i–iii) of local realism could fail, as in a standard Bell inequality. This is certainly possible in our case since the Bell qubits are neighbors on the same superconducting chip. However, arranging for a locally realist model that accounts for the needed disturbance effects for the neighboring Bell qubits, the neighboring Bell-ancilla qubits, each remote pair of Bell-ancilla

qubits, and the remote ancilla-ancilla qubits simultaneously is substantially more difficult (and therefore much less likely) than arranging for such disturbance in the usual Bell test on just two neighboring qubits. Moreover, our experiment verifies the detailed functional dependence of the quantum predictions as the weak measurement angle $\phi$ is varied, which further constrains any purported locally realist explanation. Thus our tested BLGI significantly tightens the locality loophole [9] compared to the usual Bell test performed on the same chip.

Second, the noisy detector assumption (iv) could fail due to hidden preparation noise $\xi_P$ not included in the state $\lambda$ that systematically affects the detector output in both arms in a correlated way. In this case, the detector response would become noise-dependent $P_A(\xi|\lambda) \rightarrow P_A(\xi|\lambda, \xi_P)$ such that the calibration of Eq. (5) will be satisfied only after additionally averaging over $\xi_P$. Such correlated noise would prevent the detector distributions from factoring for each $\lambda$ in Eq. (7), which formally spoils the inequality. However, in our experiment such a systematic bias due to correlated noise has been extensively checked during the characterization of the chip and the measurement calibration by deliberately preparing a variety of uncorrelated distributions $P(\lambda)$ (i.e., different qubit states) and looking for spurious cross-correlations of the various qubit readout signals that would be expected in the presence of such hidden preparation noise. Hence, the failure of assumption (iv) additionally requires an unlikely preparation-conspiracy where every calibration check that has been done is somehow immune to the hidden detector-noise correlations.

**Weak Measurement Calibration**

As discussed in the main text the ancilla readout is imperfectly correlated to the Bell qubit's state. When measuring in the $Z$ basis, $\langle Z \rangle_\alpha = \sin(\phi)\langle Z \rangle_\beta$, shown in Fig. 2 (a) along with the ideal curves $\pm\sin(\phi)$. To calibrate this weak measurement we must first relate the measurement angle $\phi$ to microwave drive power, by fitting to a measurement of $|1\rangle$ state probability vs. $\pi$-pulse amplitude. The most straight forward calibration would then be to divide $\langle Z \rangle_\alpha$ by $\sin(\phi)$ shown in the blue curves in Fig. 2 (b), but this method causes drift in the mean at the smallest angles.

This simple calibration fails because the raw data curves shown in Fig. 2 (a) converge to a value slightly below zero. This means that for the weakest measurements, the simple angle calibration will under-correct a $|1\rangle$ state measurement and over-correct a $|0\rangle$ state measurement. This over-correction of the $|0\rangle$ state is problematic, since calibrated values for $\langle Z \rangle_\alpha$ not bounded by $\pm 1$ will possibly violate the inequality incorrectly. To prevent this, we instead use a data based calibration

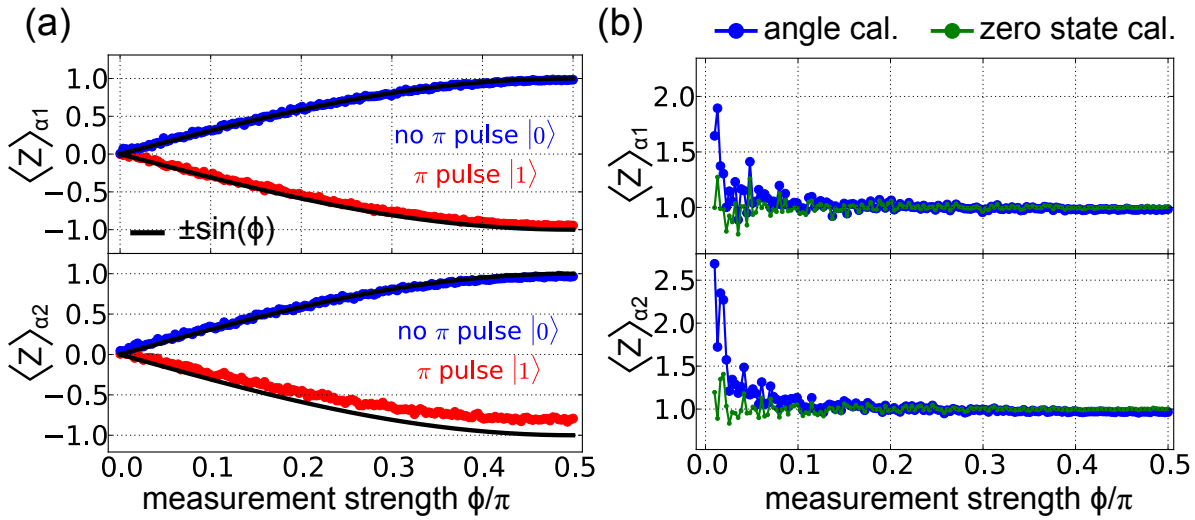


Figure 2. (a) Raw measurement of $\langle Z \rangle$ vs. measurement strength $\phi$ for both $\alpha_1$ and $\alpha_2$ along with an ideal curve of $\pm \sin(\phi)$. Due to decoherence, measurement error, and possible calibration error the data does not perfectly follow the $\pm \sin(\phi)$ curves and converges to a value slightly below 0 at the weakest measurement strength. (b) $|0\rangle$ state calibrations done using both the $1/\sin(\phi)$ method (blue) and the point by point calibration (green). In the $1/\sin(\phi)$ calibration the mean of the $|0\rangle$ state measurement drifts above 1 at the weakest measurement angles, while the mean follows 1 for the point by point case.

for each ancilla using the average of the $|0\rangle$ state measurement curve. This has the advantage of bounding the mean of the calibrated result by $\pm 1$, at the expense of accentuating the drift in the mean of the $|1\rangle$ state measurement towards 0. The results of this $|0\rangle$ state calibration are shown in the green curves in Fig. 2 (b).

To apply this calibration to the correlator terms, we must first express them in terms of the measurement operator $\langle Z \rangle$. In a superconducting system the state rotations are used to map the desired measurement basis onto the ground ($|0\rangle$) and excited ($|1\rangle$). For the ancilla measurement this is equivalent to mapping onto the $Z$ measurement axis. Given probability P(1) of measuring the excited ($|1\rangle$) state, $\langle Z \rangle = 1 - 2P(1)$. After mapping state probabilities onto the $Z$ measurement axis we can express the correlator as $E(\alpha, \beta) = \langle Z \rangle_\alpha \langle Z \rangle_\beta$. Expressed in this way we can see that for calibration factor $cal(\phi) \approx 1/\sin(\phi)$, $E_{cal}(\alpha, \beta) = E(\alpha, \beta) * cal(\phi)$. Extending this to the BLGI we calibrate each term depending on the ancilla qubit being measured such that $E(\alpha_1, \beta_2) \rightarrow E(\alpha_1, \beta_2) * cal(\phi_1)$, $E(\beta_1, \alpha_2) \rightarrow E(\beta_1, \alpha_2) * cal(\phi_2)$, $E(\alpha_1, \alpha_2) \rightarrow E(\alpha_1, \alpha_2) * cal(\phi_1) * cal(\phi_2)$, and $E(\beta_1, \beta_2)$ remains unchanged.

**Error Analysis and Pulse Sequence Optimization**

While the algorithm and weak measurement scheme are simple in design, dependence on correlations between multiple qubits makes $\langle C \rangle$ sensitive to multiple error mechanisms. As all four qubits were operated away from the flux insensitive point they were more susceptible to dephasing effects. The Xmon qubits used in this experiment were extensively characterized in Ref. [10], with a $T_2$ Ramsey decay of order $2 - 3\,\mu$s at the idle point. When Characterizing the phase error per gate with randomized benchmarking, we find an error of roughly 0.25 percent per gate. This corresponds to a $T_{2\text{echo}}$ of over $10\,\mu$s. The amplitude of $\langle C \rangle$ vs. dephasing error per qubit is shown in Fig. 3(a). The violation amplitude is relatively robust to this error, and can sustain error rates of up to 30 percent while still exhibiting non-classical correlations. The second major error mechanism was reduced measurement visibility coming from $T_1$ energy decay or spurious $|1\rangle$ state population. The effect on $\langle C \rangle$ vs. single qubit measurement visibility is shown in Fig 3(b). The correlation amplitude is more sensitive to this reduced measurement visibility and is significantly degraded at even 90 percent. In both cases, the presence of errors not only lowers the maximum vi-

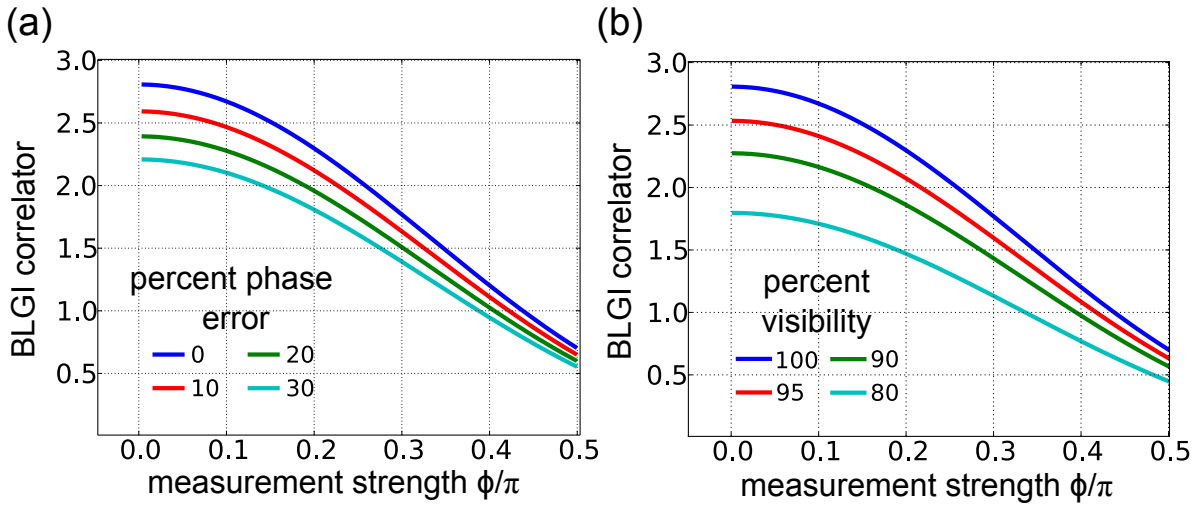


Figure 3. Effect of error mechanisms on BLGI correlations. (a) Amplitude of $\langle C \rangle$ vs. measurement strength for various single qubit dephasing error rates. The single qubit dephasing error rate can be thought of as roughly $T_2$/(gate time) in the simplest case, but can be improved through the addition of echo pulses. (b) Amplitude of $\langle C \rangle$ vs. measurement strength for various single qubit measurement visibility values. The system shows a greater sensitivity to this error mechanism and cannot tolerate a visibility much below 90 percent.

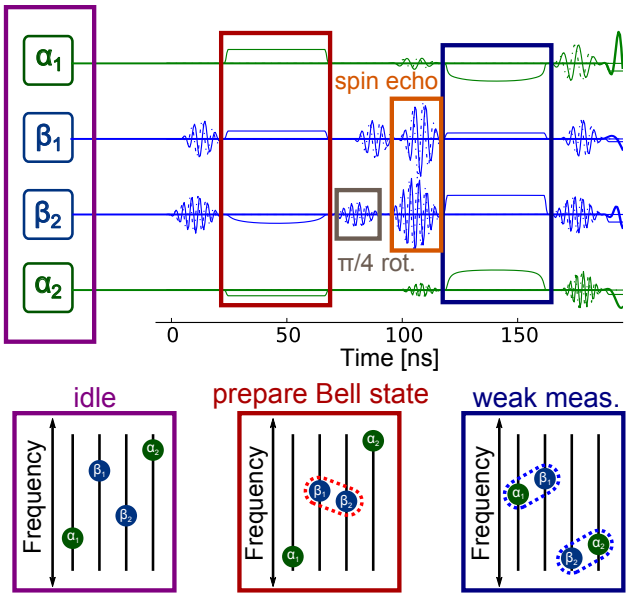


Figure 4. (above) Full pulse sequence of the BLGI algorithm, including spin echo pulses on the Bell qubits to reduce dephasing. The weak measurements were carried out simultaneously to further streamline the algorithm. (below) Frequency diagram of the 4 qubits showing their placement in frequency space the pulse sequence, in particular the various detunings during the adiabatic $CZ$ gates.

olation possible but the highest measurement strength at which a violation first occurs. As weaker measurement angles require finer calibration and provide noisier data, it is preferable to achieve a violation at the largest measurement strength possible.

Given the sensitivity of $\langle C \rangle$ to various decoherence mechanisms it was important to reduce the BLGI pulse sequence time as much as possible for higher coherence. This is most notable during the weak measurement portion when we carry out simultaneous adiabatic CZ gates[11] between both ancilla-Bell pairs. Lastly, we introduced spin echo pulses in the middle of the algorithm which cancel out dephasing during the pulse sequence while simply transforming the original $|\Psi^+\rangle$ Bell state to a $|\Psi^-\rangle$. To maximize measurement fidelity, we used a wide bandwidth parametric amplifier [12], to ensure a high signal to noise ratio and shorter readout time. A separate measurement at the beginning of the pulse sequence was used to herald [13] the qubits to the ground state, but this was a small ($\sim$6 percent) effect. Lastly, we implemented numeric optimization of the adiabatic $CZ$ gates using the ORBIT protocol [14] to fine tune parameters for the final data set. The full Pulse sequence and frequency placement of the qubits during the algorithm is shown in Fig. 4.

During the numeric optimization of the pulse sequence, single qubit phases can be adjusted

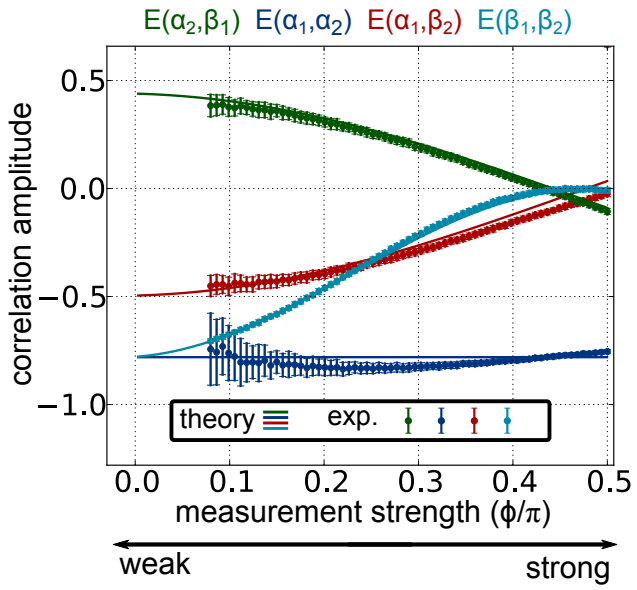


Figure 5. Graph showing both experimental data (points) and theoretical predictions (lines) for the four BLGI terms vs. measurement strength $\phi$. The data set was taken by averaging together 200 hundred traces in which each point was measured 3000 times for a total of 600,000 iterations per point. The error bars represent 10 standard deviations of the mean to demonstrate the scaling the ancilla measurement noise vs. measurement strength. Distinct differences in the behavior arise from experimental imperfections and their depedence on different qubit correlations. In particular the difference between $E(\beta_1, \beta_2)$ and $E(\alpha_1, \beta_2)$ or $E(\beta_1, \alpha_2)$ highlights the dependence of entanglement collapse on the measurement strength of both ancilla qubits.

slightly to increase correlations, leading to a larger violation. This is equivalent to changing the final rotation angle of the detectors slightly ($\sim$ 3 degrees). The nature of the BLGI makes it immune to such rotations as loss of correlations in one correlator is naturally made up for in another. additionally, the initial detector rotation $b$ was chosen based on the maximum of the original CHSH measurements. Due to differences in qubit coherence this does not necessarily occur at $\pi/4$, but at a slightly smaller angle. The individual BLGI correlator terms measured in this experiment along with theory curves accounting for these realistic rotations are plotted in Fig. 5. The behavior of each individual term depends on the type of qubits being correlated. The term $E(\alpha_1, \alpha_2)$ holds roughly constant, close to the expected value of $-1/\sqrt{2}$. $E(\alpha_1, \beta_2)$ and $E(\beta_1, \alpha_2)$ start close to zero, and converge to around $\pm 0.5$. The behavior of $E(\beta_1, \beta_2)$ best matches expectations. It begins at 0 for strong ancilla measurement and converges near $1/\sqrt{2}$ at

perfectly weak measurement, following a slightly s-shaped curve.

### Sample fabrication and characterization

Devices are fabricated identically to Ref. [11], and extensive calibration was documented in Ref. [10].

### Device parameters

The device parameters are listed in table I. Note that the coupling rate $g$ is defined such that strength of the level splitting on resonance (swap rate) is $2g$ (Ref. [15]).

Table I. Parameters for the device when running the BLGI algorithm. $f$ are frequencies. $\eta$ is qubit nonlinearity. $g$ is coupling strength. $\kappa$ is resonator leakage rate.

| | $Q_0$ | $Q_1$ | $Q_2$ | $Q_3$ |
|---|---|---|---|---|
| **Qubit frequencies and coupling strengths** | | | | |
| $f_{10}^{max}$ (GHz) | 5.30 | 5.93 | 5.39 | 5.90 |
| $\eta/2\pi$ (GHz) | -0.230 | -0.216 | -0.229 | -0.214 |
| $f_{10}^{idle}$ (GHz) | 4.53 | 5.42 | 4.67 | 5.55 |
| $f_{\mathrm{res}}$ (GHz) | 6.748 | 6.626 | 6.778 | 6.658 |
| $g_{\mathrm{res}}/2\pi$ (GHz) | 0.110 | 0.128 | 0.111 | 0.109 |
| $g_{\mathrm{qubit}}/2\pi$ (MHz) | 13.8 | | 14.1 | |
| $g_{\mathrm{qubit}}/2\pi$ (MHz) | 14.5 | | | |
| $1/\kappa_{\mathrm{res}}$ (ns) | 675 | 69 | 555 | 30 |
| **Readout (RO) parameters** | | | | |
| RO error | 0.015 | 0.004 | 0.067 | 0.007 |
| Thermal $\|1\rangle$ pop. | 0.013 | 0.007 | 0.028 | 0.01 |
| RO pulse length (ns) | 1000 | 300 | 1000 | 300 |
| RO demodulation length (ns) | 1000 | 300 | 1000 | 300 |
| **Qubit lifetime at idling point** | | | | |
| $T_1$ ($\mu$s) | 26.3 | 24.7 | 39.2 | 21.3 |


[1] Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880 (1969).

[2] Bell, J. S. *et al.* On the einstein-podolsky-rosen paradox. *Physics* **1**, 195–200 (1964).

[3] Emary, C., Lambert, N. & Nori, F. Leggett–garg inequalities. *Rep. Prog. Phys.* **77**, 016001 (2014).

[4] Wilde, M. M. & Mizel, A. Addressing the clumsiness loophole in a Leggett-Garg test of macrorealism. *Foundations of Physics* **42**, 256–265 (2012).

[5] Knee, G. C. *et al.* Violation of a Leggett-Garg inequality with ideal non-invasive measurements. *Nat. Comm.* **3**, 606 (2012).

[6] Palacios-Laloy, A. *et al.* Experimental violation of a Bell/'s inequality in time with weak measurement. *Nat. Phys.* **6**, 442–447 (2010).

[7] Groen, J. *et al.* Partial-measurement backaction and nonclassical weak values in a superconducting circuit. *Phys. Rev. Lett.* **111**, 090506 (2013).

[8] Larsson, J.-Å. Bell's inequality and detector inefficiency. *Phys. Rev. A* **57**, 3304 (1998).

[9] Larsson, J.-Å. Loopholes in Bell inequality tests of local realism. *J. Phys. A* **47**, 424003 (2014).

[10] Kelly, J. *et al.* State preservation by repetitive error detection in a superconducting quantum circuit. *Nature* **519**, 66–69 (2015).

[11] Barends, R. *et al.* Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature* **508**, 500–503 (2014).

[12] Mutus, J. *et al.* Strong environmental coupling in a Josephson parametric amplifier. *Appl. Phys. Lett.* **104**, 263513 (2014).

[13] Johnson, J. *et al.* Heralded state preparation in a superconducting qubit. *Phys. Rev. Lett.* **109**, 050506 (2012).

[14] Kelly, J. *et al.* Optimal quantum control using randomized benchmarking. *Physical review letters* **112**, 240504 (2014).

[15] In refs. [11] and [14], the convention was used where $g/2\pi$ is the swap rate. Here we use $2g/2\pi$ as the swap rate, in accordance with others in the field.